ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/183880/

This is the author’s version of a work that was submitted to / accepted for publication.

Citation for final published version:

Zhang, Haiyang, Yan, Wei, Chen, Chong, Liu, Ying , Liu, Qingtao, Liang, Xiaolei and Jiang, Zhigang 2026. A physics-informed multimodal transformer model for machining energy consumption prediction. Journal of Computing and Information Science in Engineering 10.1115/1.4070798

Publishers page: http://dx.doi.org/10.1115/1.4070798

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

A physics-informed multimodal transformer model for machining energy consumption prediction

Haiyang Zhang, first author
Affiliation: School of Automobile and Traffic Engineering, Wuhan University of Science and Technology, Wuhan, 430081, China.
E-mail: ZHANGHY678678@163.com
Wei Yan, second author
Affiliation: School of Automobile and Traffic Engineering, Wuhan University of Science and Technology, Wuhan, 430081, China.
E-mail: yanwei81@wust.edu.cn
Chong Chen, third author
Affiliation: Guangdong Provincial Key Laboratory of Cyber-Physical System, Guangdong University of Technology, Guangzhou, 510006, China.
E-mail: chenc2021@gdut.edu.cn
Ying Liu, fourth author
Affiliation: Department of Mechanical Engineering, School of Engineering, Cardiff University, Cardiff CF24 3AA, UK.
E-mail: liuy81@cardiff.ac.uk
Qingtao Liu, fifth author
Affiliation: Key Laboratory of Road Construction Technology and Equipment of MoE, Chang'an University, Xi'an, 710064, China.
E-mail: qtaoliu@chd.edu.cn
Xiaolei Liang, sixth author[1]
Affiliation: School of Automobile and Traffic Engineering, Wuhan University of Science and Technology, Wuhan, 430081, China.
E-mail: liangxiaolei@wust.edu.cn
Zhigang Jiang, seventh author
Affiliation: Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan, 430081, China.
E-mail: jiangzhigang@wust.edu.cn

Abstract

Machining energy consumption (MEC) prediction plays an important role in energy planning, management, and conservation in the manufacturing industry. Existing MEC prediction techniques are commonly classified into physics-based models (PBMs) and data-driven approaches (DDAs), which rely on physical knowledge and data learning, respectively. While DDAs alleviate limitations associated with PBMs due to assumptions needed to simplify the complexity, most of them ignore the underlying physical knowledge and may struggle with generalization under varying machining conditions. Moreover, machining data is inherently heterogeneous in modalities, such as sensor signals and process parameters, making data fusion another challenge. To address the above issues, this paper proposes a physics-informed multimodal transformer (PIMT) model for MEC prediction. Firstly, based on the studies of machining energy nature, a synthetic data augmentation strategy is

[1] Corresponding author.

proposed to incorporate energy-relevant physical principles from PBMs. Secondly, a structured multimodal network is designed to align and fuse features from diverse data modalities. Then, a transformer with multi-head attention is employed to capture complex temporal dependencies and cross-modal interactions within the fused data. Finally, three groups of comparative experiments with different machine tools, materials, cutting tools, and machining parameters are designed to validate and demonstrate the proposed approaches. The results showed that the proposed PIMT model achieves the lowest RMSE (0.048) and MAPE (12.33%), while reducing the training time (at least 152s, 1.7s, and 29.9s) and the number of trainable parameters (at least 49.7%, 51.3% and 67.6%) compared to state-of-the-art methods. Furthermore, the PIMT model consistently produces smaller errors across nearly all evaluation metrics, and accurately forecasts MEC throughout continuous machining conditions.

Keywords: Physics-informed machine learning; Multimodal data fusion; Transformer; Energy consumption prediction; Machining.

1. Introduction

Global energy consumption exerts a profound impact on both the environment and national economies [1]. Within manufacturing sectors, machining is a major contributor to carbon dioxide emissions and energy usage, responsible for over 74.7% of global energy consumption and 99% of carbon footprint [2-3]. It is a strong motivation to reduce machining energy consumption (MEC). Thus, various strategies, such as process parameter optimization and energy-saving scheduling, have been proposed to decrease the energy consumption and carbon dioxide emissions of machining [4]. However, the absence of an accurate MEC model has hindered the effective implementation of these strategies [5].

Existing MEC prediction techniques are commonly classified into physics-based models (PBMs) and data-driven approaches (DDAs). PBMs primarily rely on understanding energy transmission and conversion laws in machining, incorporating principles of thermodynamics [6], or energy flow theory [7]. One limitation in PBMs is the model fidelity, as the physical parameters in equations may only be roughly approximated due to the non-linearity involved, which will limit the prediction accuracy. With the development of machining learning and deep learning technologies, DDAs gain attention as a viable alternative to resolve this issue [8]. DDAs do not assume physical knowledge about the system to be available a priori. Instead, they establish numerical mappings that directly associate sensing data to system variables of interest by learning from the training samples [9]. However, DDAs often act as a type of "black box" where the inputs and outputs are known, but the underlying process for deriving the relationships among these remains a mystery [10]. Moreover, machining conditions are often dynamically changing (such as variations in machine tools, machining processes, and workpieces), which introduces inherent uncertainties in MEC prediction and may also lead to potential inaccuracies in the results of DDAs. These limitations of both PBMs and DDAs highlight the need for an approach that integrates physical principles with data-driven learning. To bridge this gap, the emerging paradigm of Physics-Informed Machine Learning (PIML) offers a promising path forward. PIML addresses these shortcomings by incorporating physical prior knowledge, which can be achieved by modifying various components of the machine learning pipeline, including the data, model architecture, or loss function, etc [11]. Therefore, a key task within this paradigm is to explore effective strategies for integrating these inherent physical prior knowledge into DDAs.

Paper number: JCISE-25-1455. Corresponding author: Xiaolei Liang

Moreover, the machining data is inherently heterogeneous in modalities. For instance, the information of machine tools, cutting tools, and workpieces, such as machine type, tool diameter, and workpiece material is typically used to describe the entities involved in machining and usually remain unchanged during machining, thus representing static data. Process parameters, such as cutting speed, feed rate, and cut depth are often collected from numerical control (NC) codes or manual operator input, describing the "intent" and "instructions" of the machining process, and are typically discrete event sequence data. Sensor data, such as spindle motor power, feed-axis power, and three-directional cutting forces, are generally high-frequency, continuous time-sequence data. Additionally, the machining condition data, such as workshop temperature and coolant flow rate, are typically structured data. These multimodal data provide a unique perspective for understanding the nature of energy consumption in machining, and their effective utilization may enhance the prediction accuracy and generalization of MEC models. However, existing research mostly focuses on single-modality inputs or simple early/late fusion strategies. Each modality exhibits unique data patterns, making it difficult to transfer models trained on one modality to another [12].

To address these challenges, this paper focuses on the data level of PIML and employs PBMs to generate synthetic features (i.e., synthetic data) that encapsulate underlying physical principles. This synthetic data acts as a physical prior, guiding the model to learn representations consistent with the known physics. These synthetic data are then used as additional input features, aligned with the other modalities for MEC prediction. Consequently, we propose a physics-informed multimodal transformer (PIMT) model. The intellectual contributions of this paper can be summarized as follows: (a) a strategy for exploring and integrating the machining physical knowledge in DDAs based on data-level PIML paradigm is proposed, enhancing the accuracy and physical consistency of the MEC model, and (b) a multi-channel transformer is designed to process the multimodal data for improving the model's generalization capability, providing a systematic and effective solution to the challenge of fusing heterogeneous data (static, sequential, and structured) in manufacturing.

The rest of this paper is organized as follows. Section 2 reviewed the relevant studies of the PBMs and DDAs, as well as multi-modal data fusion for MEC modelling and prediction. Section 3 introduces the methodology of the proposed PIMT model. Section 4 reports a series of experimental studies, and the results are demonstrated and discussed in Section 5. Finally, conclusions were presented in Section 6 to summarize our contributions and to outline the need for future research.

2. Literature Review

2.1. MEC modelling and prediction

MEC prediction can be primarily divided into PBMs and DDAs. PBMs rely on understanding of physical principles or empirical formulas in machining, such as machining mechanisms and energy rules. Gutowski et al. presented an exergy-based theoretical model for MEC [6]. Subsequently, other researchers conducted studies to develop empirical models to predict energy consumption in machining processes, including turning, milling, drilling, etc. [13]. Liu et al. developed power balance equations of machine tool steady and transient periods based on the energy flow theory [14]. Similarly, Mori et al. established an energy consumption model that considers both cutting and non-cutting periods of machine tools, with a specific focus on energy consumption resulting from spindle positioning and acceleration [15]. Balogun and Mativenga established a predictive model that accounts for energy consumption in the basic state, ready state, and cutting state, offering valuable insights for managing the energy demand of machine tools [16]. Ma et al. proposed a material removal energy consumption model during milling, considering the cutting and air-cutting status of machine tools [17]. Liu et al. extended the scope of their study by developing a model to estimate MEC, including both the direct and indirect energy consumption associated with the embodied energy of cutting fluid, cutting tool, and workpiece [18]. With further advancements in research, more precise energy consumption PBMs for machine tools have been developed.

In contrast to PBMs, DDAs leverage machine learning techniques to establish numerical mappings directly correlating collected data with MEC. The performance of DDAs is mainly affected by the quality of raw data, features used as model inputs and prediction techniques used for model establishment [19]. In most of the DDAs, sensor data, process parameter, equipment information, and machining condition data are commonly used. Hu et al. established an energy consumption model during cutting based on feature technology [20]. Brillinger et al. [21] investigated the ability of machine learning algorithms for MEC prediction, e.g., Decision Tree, Random Forest, and boosted Random Forest. Bhinge et al. [22] developed a Gaussian regression-based data-driven model for predicting milling energy consumption. Sheng et al. proposed a self-adaptive hybrid approach for machining quality prediction with multimodal sensing data in the manufacturing workshop [23]. More recently, the application of deep learning has emerged in this domain, capitalizing on its advantages in feature extraction and large-scale data processing. Techniques such as one-dimensional Convolutional Neural Networks (1D-CNN) [24] and deep learning-embedded semi-supervised learning (DLeSSL) [25] are gaining traction. This evolution underscores the burgeoning potential of DDAs in MEC forecasting, fuelled by the enrichment of data and advancements in data processing technologies. Recently, deep learning has significantly improved the performance of DDAs by enabling them in massive data processing and complex high-level feature extraction [26]. Particularly, the Transformer, proposed by Vaswani et al. in 2017 [27], has made a significant difference in deep learning. The Transformer eliminates the constraint of sequential processing and enables parallel processing. Moreover, benefitting from the multi-head self-attention mechanism, both long-term and short-term dependencies of sequential data could be captured effectively. In our previous work, the Transformer was employed to diagnose compound faults of industrial robots [28], further highlighting its potential utility in MEC modelling and prediction.

DDAs have made significant progress on MEC prediction. However, existing DDAs for MEC prediction primarily focus on feature extraction from machining data, failing to fully exploit the underlying physical knowledge behind these data. To this end, researchers have recently turned to the paradigm of PIML, which provides a systematic framework for integrating physical prior knowledge into models. To be specific, the training of a machine learning model involves several fundamental components including data, model architecture, loss functions, etc. The incorporation of physical prior knowledge can be achieved through modifications to one or more of these components. For example, Liu et al. [29] suggested a hybrid approach that combines data-driven machine learning with machining mechanics for specific cutting energy (SCE) modelling. Lv et al. [30] proposed a comprehensive framework to integrate both the mechanical and data aspects of MEC, and it provided valuable insights for energy consumption evaluation and management. Simultaneously, several efforts have been made to develop physics-informed modelling techniques in tool remain life prediction and heat transfer modelling, aiming to improve the performance of DDAs [31]. These studies offer practical physically consistent solutions by introducing appropriate observational, inductive, or learning biases that can steer the learning process. To the best of our knowledge, the application of these methods in MEC modelling and prediction has not been explored in the existing literature.

2.2. Multimodal data fusion

Multi-modal data fusion is defined as a framework that combines data from diverse sources and formats, and aims to achieve enhanced accuracy and reliability of inferences compared to individual data sources [32]. This technique has gained significant traction in manufacturing industries, including aerospace, automotive, and additive manufacturing, leveraging advanced sensing and Internet of Things technologies [33]. Existing methods for multi-modal data fusion can be broadly categorized into model-independent and model-based approaches. Model-independent methods, such as early, intermediate, and late integration, have been successfully applied across various domains [34]. However, the relative data for MEC modelling, such as machine tool type, workpiece material, and cutting tool type, can be acquired from process manuals, CAD models and CAM systems, while other data require the use of auxiliary devices, such as ultrasonic sensors, AE sensor, current sensor, and similar devices. Moreover, in some physics-informed neural networks, the data used for DDAs is derived from empirical equations of PBMs [35]. Therefore, directly applying these methods to integrate these multimodal data presents challenges.

Model-based data integration approaches have become popular in multi-modal data fusion due to their high accuracy and usefulness [36]. The data utilized for MEC are obtained from various information sources and presented in different formats, leading to heterogeneity in terms of their characteristics [37]. Specifically, these data formats include static data, discrete event sequence data, continuous time-sequence data, and structured data. Various methods are commonly used in this context, including multi-kernel learning, image modelling, and neural networks. Among these methods, deep learning, including various sub-neural networks, is skilled in multi-modal data processing [38], such as static data, structured data, image data, and sequential data, respectively. Zamiela et al. proposed a thermal physics-informed PointNet methodology and fused multi-modal position, weld parameters, and thermal data to monitor the process for defects [39]. Chen et al. developed a dual-branch deep neural network to fuse satellite and street-view images [40]. Mou et al. [41] established a CNN-LSTM model to fuse multi-modal data, including eye data, vehicle data, and environmental data. Consequently, multi-modal data fusion can therefore play an essential role in understanding the energy nature of machine tools. The development of deep learning has improved the ability for multi-modal data fusion, and it has generated great interest in multimodal data processing and general concern among researchers.

From the literature, MEC modelling and prediction have been developed based on both PBMs and DDAs. PBMs provide a theoretical foundation for understanding the energy nature of machine tools, while DDAs offer flexibility and adaptability for modelling. However, a fundamental challenge lies in exploiting the physical knowledge of MEC. Research on PBMs of MEC has yielded significant results, but existing DDAs typically employ data collected from various machining conditions, and the physical knowledge hidden in PBMs is insufficiently exploited, resulting in DDAs that are not adequately physically informed. Another important challenge arises from the complexity of data. The data related to MEC provide a multidimensional understanding of the energy nature, but the lack of integration of these multisource and heterogeneous data limits the performance of DDAs. Therefore, it is critical to explore and utilize the physical knowledge and integrate diverse sources and heterogeneous information to gain a comprehensive understanding of MEC.

3. Methodology

The proposed PIMT model for MEC prediction is grounded in the data level of the PIML paradigm. This approach integrates physical knowledge by generating synthetic features from PBMs. These features serve as physical priors that enrich the input space and guide the model toward physically consistent representations. As illustrated in Fig. 1, the overall framework of the approach consists of two core modules: physical knowledge exploration and PIMT model for MEC prediction.

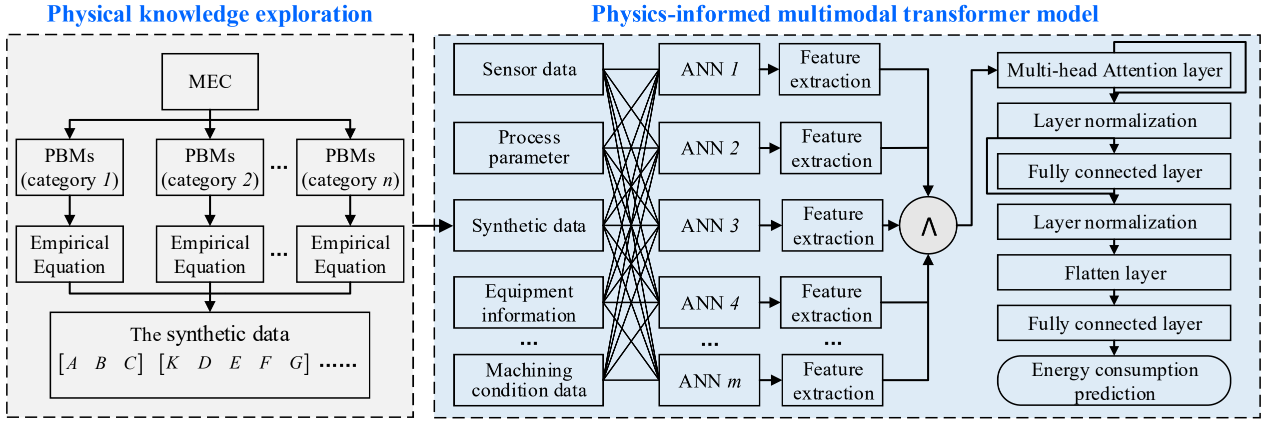

Figure 1. The general framework of the PIMT model for MEC prediction

(1)The physical knowledge exploration module is utilized to discover the latent physical insights within PBMs. In this study, we employ an exploration strategy to generate synthetic data characterizing physical knowledge. To initiate this process, PBMs for total MEC were formulated including different energy categories, mainly considering air cutting energy consumption and cutting energy consumption. Subsequently, synthetic data were generated using empirical equations of each category.

(2)The PIMT model module is established to integrate multi-modal data using a hybrid neural network and predict MEC with a transformer. Firstly, suitable deep learning networks are designed to extract the high-level features from synthetic data and other data (sensor data, process parameters, equipment information, and machining condition data). Alignment and merging strategies are then applied to fuse these features. Second, the fused features are fed into a transformer with a multi-head self-attention mechanism for MEC prediction. The model takes the five aforementioned data streams as input and outputs the real-time energy consumption.

3.1. Physical knowledge exploration

In general, the energy consumption during CNC machining with complicated dynamic changes, exhibiting a multi-category energy nature. Specifically, the energy consumption can be divided into mainly four categories, i.e., standby, air cutting, cutting, tool change and external auxiliary. A machine tool power profile during a typical face milling process is shown in Fig. 2.

Throughout the machining process, standby time is strongly affected by the operator's skill level. Because operators generally lack energy-saving awareness, the machine's runtime under the standby state tends to fluctuate randomly. To simplify calculations, the energy consumed during this state is excluded. Moreover, the energy required for spindle and feed-axis acceleration and deceleration is negligible, as these transient events are extremely brief. Consequently, the total energy consumption E_{total} of the entire machining process can be calculated directly using Equation (1).

Paper number: JCISE-25-1455. Corresponding author: Xiaolei Liang

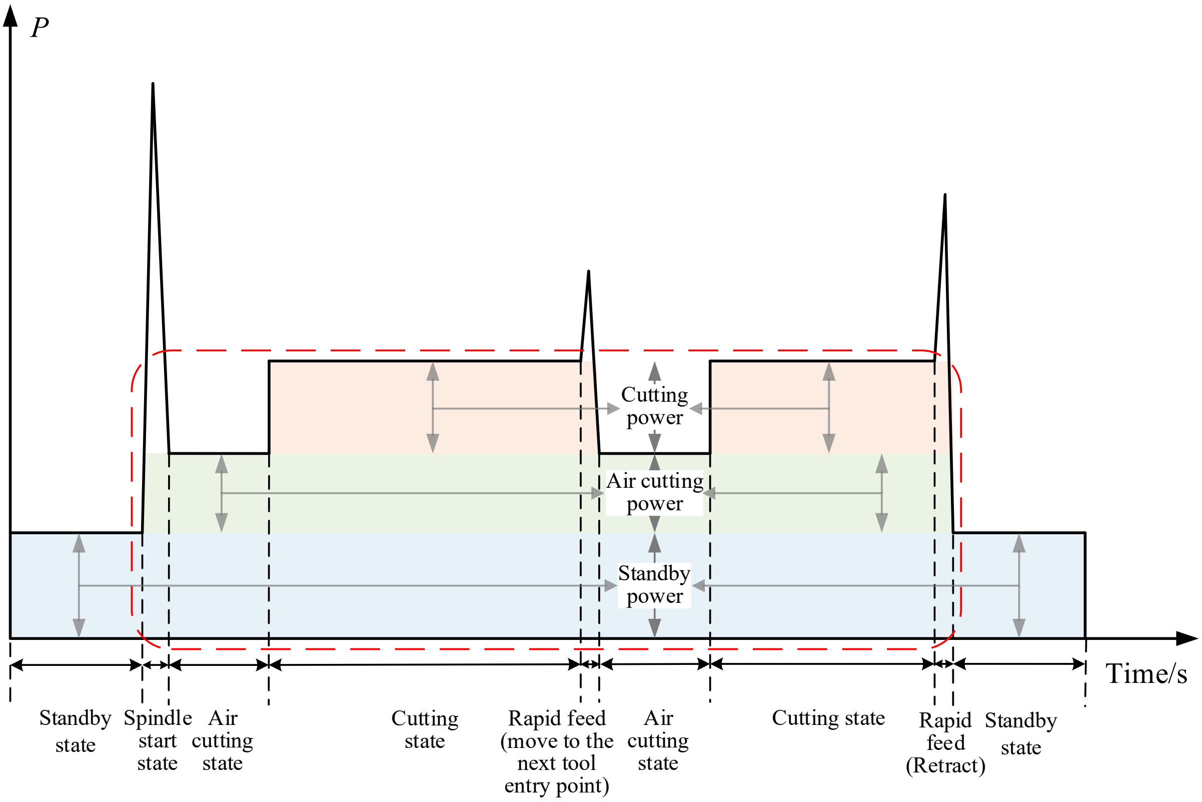

Figure 2. Power profile of machine tool during the machining process

$$E_{total} = E_{air} + E_{cut} + E_{aux} = \int^{T_1} P_{air}(t)dt + \int^{T_2} P_{cut}(t)dt + \sum_{j=1}^{n} S_i(t)E_j \tag{1}$$

where E_{total} represents the total energy consumption of the entire machining process. E_{air}, E_{cut}, and E_{aux}, denotes energy consumption in air cutting, cutting, and auxiliary state, respectively; P_{air}, P_{cut}, and. T_1, T_2 are the power and duration corresponding to these stages, respectively; E_j is the energy consumption of the auxiliary subsystem $(j = 0,1,2,\cdots n)$, $S_j(t)$ is the switch function as shown in Eq. (2), indicating whether the auxiliary subsystem is on or off.

$$S_j(t) = \begin{cases} 1 \ systems \ on \\ 0 \ systems \ off \end{cases} \tag{2}$$

Considering the linear accumulation characteristics of auxiliary system energy consumption on a macroscopic time scale, this study simplifies it into a constant C based on the empirically measured average value during auxiliary system operation, thereby striking a balance between model complexity and engineering practicality.

During face milling, the cutter and the workpiece rotate continuously relative to each other while the table feeds radially at a constant rate. Because the spindle speed and the feed motion are coupled through a fixed transmission ratio, the instantaneous power of the spindle motor can be regarded as a single-factor function of the cutter speed. Moreover, variations in the radial feed rate have a negligible effect on the feed-axis motor power, so this contribution can be treated as constant. Consequently, the air cutting power of the milling machine can be simplified to a function of the cutter speed. Experimental fitting results and published studies both indicate a quadratic relationship between total motor power and spindle speed [17]. Therefore, the air cutting power P_{air} is expressed as Equation (3):

$$P_{air}(n) = An(t)^2 + Bn(t) + C \tag{3}$$

where $n(t)$ represents the spindle rotation speed of machine tools; A, B, and C are physical parameters.

In the cutting state, material removal is accomplished through the combined action of the spindle and feed axes, achieved by changing the tool position. Therefore, the cutting power P_{cut} is expressed as Equation (4).

$$P_{cut}(n, a_{sp}, f, a_{se}) = K \cdot n(t)^D \cdot a_{sp}^E \cdot f^F \cdot a_{se}^G \tag{4}$$

where $n(t)$, a_{sp}, f, and a_{se} represent the spindle rotation speed, depth of cut, feed rate, and width of cut, respectively; K, D, E, F, and G are physical parameters.

The multi-categories Eqs. (3)–(4) express machining power as a function of real-time process parameters (e.g., spindle speed, feed rate). Then, the physical parameters (i.e., synthetic data) corresponding to a specific machining state are fitted using empirical equations, with real-time measured process parameters (e.g., spindle speed and feed rate) as inputs and corresponding machining powers (i.e., P_{air}, P_c) as outputs. It is worth noting that the aforementioned process parameters and corresponding powers are grouped under the same machining state. Finally, all synthetic data are used as additional input features, which are aligned with the other modalities for MEC prediction. The symbols for synthetic data of every machining state acquisition are shown in Table 1.

Table 1. The synthetic data acquisition

Machining state	Energy category	Empirical equation	Input	Synthetic data
State 1	Air cutting	$P_{air} = An^2 + Bn + C$	n	$[A_1 \quad B_1 \quad C_1]$
State 2				$[A_2 \quad B_2 \quad C_2]$
......				......
......	Cutting	$P_c = K \cdot n^D \cdot a_{sp}^E \cdot f^F \cdot a_{se}^G$	n, a_{sp}, f, a_{se}	......
State n				$[K_n \quad D_n \quad E_n \quad F_n \quad G_n]$

3.2. Physics-informed multimodal transformer model

In this work, the generated synthetic data, sensor data, process parameters, equipment information, and machining condition data are used for MEC prediction. From the perspective of the modal characteristics of these data, the equipment information and machining condition data are static and structured data, which formats apply to FCNN. Among these, one-hot encoding is used to convert categorical equipment information (e.g., type of machine tool, type of material) into numerical formats suitable for the FCNN. The sensor data and process parameters are continuous time-sequence data and discrete event sequence data, which formats apply to 1D-CNN. Similarly, since the synthetic data is fitted from the process parameter, its sampling frequency is the same as the process parameter, which also applies to 1D-CNN. Thus, a PIMT model including an FCNN

Paper number: JCISE-25-1455. Corresponding author: Xiaolei Liang

and a 1D-CNN is proposed to fuse these multimodal data, then a Transformer with multi-head self-attention is employed to predict the EMC, and the architecture of the model is shown in Fig. 3.

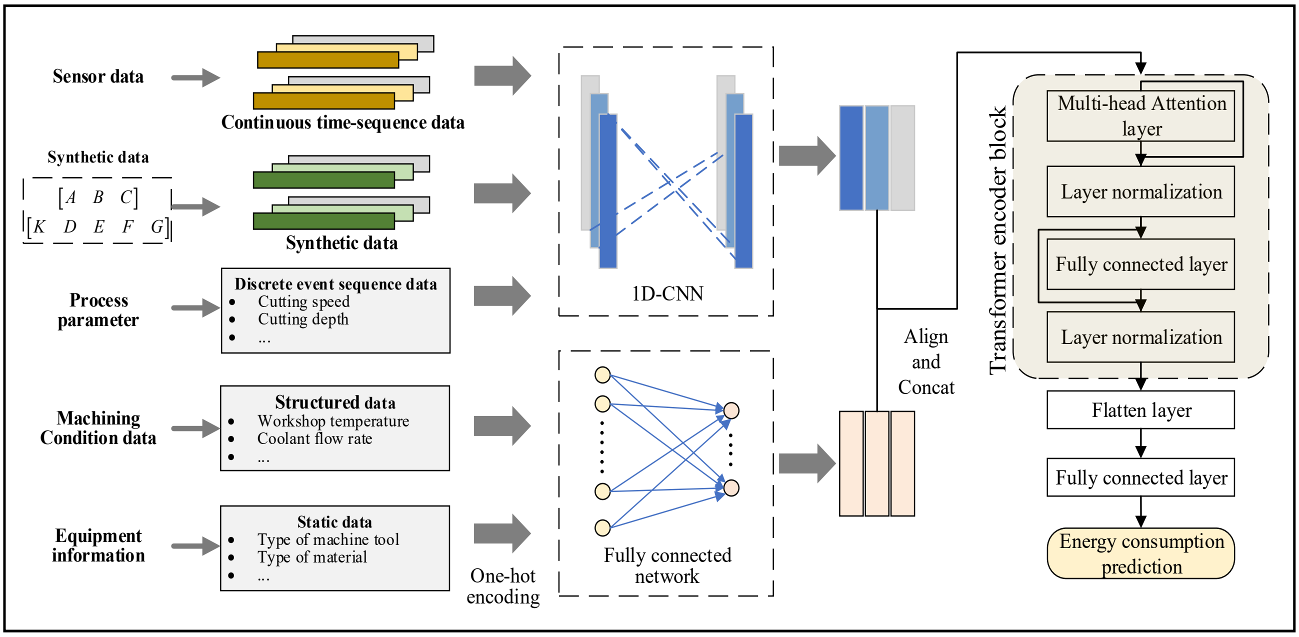

Figure 3. The architecture of the proposed PIMT model.

3.2.1 1D-CNN

A 1D-CNN is possibly the latest achievement in the era of deep learning, which is very suitable for pattern recognition in multi-channel signal processing [42]. Considering the temporal nature of sensor data, process parameter and synthetic data, a 1D-CNN is adopted to effectively integrate these sequential data. In this approach, the preprocessed sequential data samples are fed from the input layer to the convolutional layer, where the generation of feature maps constitutes the most computationally intensive task. Within each convolutional layer, the input data undergoes convolution with a kernel that has a localized receptive domain. Then, a bias term is added to generate an output feature map through a nonlinear activation function, which is fed into subsequent convolutional layers. The convolution operations can be expressed mathematically as equation (5).

$$z_i = \sigma(W_l^T x_{i,j+F_{l-1}} + b) \tag{5}$$

where $W \in R^{F_l}$ represents the convolutional kernel, b represents the bias term, $x_{i,j+F_{l-1}}$ is a subsequence of x with length F_l from point i, and z_i represents the learned feature. $\sigma(\ )$ denotes the activation function.

3.2.2 Transformer network

A transformer network can extract valuable temporal features from time series data and achieves better performance than other neural networks in various tasks. The Transformer consists of two essential components, positional encoding and a multi-head attention mechanism [43]. The extracted features are initially input into the multi-head self-attention layer to capture crucial information from diverse perspectives. To obtain the attention output, it is necessary to establish the relationship between a query and a set of key-value pairs. The attention output is a weighted sum of the values, which indicate the location of captured information. Meanwhile, the weights are computed using a compatibility function that relates the query to the corresponding key. The

Paper number: JCISE-25-1455. Corresponding author: Xiaolei Liang

standard self-attention layer calculates attention score by computing the query vector q, key vector k, and value vector v. When a new input i is fed into a self-attention layer, the attention score is calculated as follows:

$$Score = soft\max(q_i * k_i)v_i \tag{6}$$

The multi-head self-attention utilizes three matrices, Q, K, and V to replace the individual vectors q, k, and v, respectively. Within these matrixes, the information is beneficial for comprehensively assessing the importance of features from various perspectives.

For the input data $X = [x_1, x_2, \cdots, x_n]$, a linear transformation is initially applied to obtain the matrices Q, K, and V, as shown in Eqs (7)-(9).

$$Q = XW^q \tag{7}$$

$$K = XW^k \tag{8}$$

$$V = XW^v \tag{9}$$

where W^q, W^k, and W^v are trainable projection matrices.

Next, the obtained matrices Q, K, and V are used as the input for the scaled dot-product attention. The expressions of the multi-head attention mechanism are described as follows.

$$h_j = \frac{soft\max(Q_j \times K_j^T) \times V_j}{\sqrt{d}} \tag{10}$$

$$MultiHead(Q,K,V) = Concat(h_1, h_2, \cdots h_m)W_o \tag{11}$$

where d is a scalable factor, h_j denotes the j th scaled dot-product attention mechanism, m is the number of heads, and W_o is a trainable weighted matrix. The output of the multi-head self-attention is then concatenated with its input and processed by the normalization layer to avoid overfitting, which can be expressed as follows.

$$y_{norm} = LayerNorm\left[X + MultiHead(X)\right] \tag{12}$$

where X is the input to the multi-head self-attention layer.

Subsequently, the extracted features y_{norm} are sent into two feed-forward layers with a skip connection. Finally, another layer-normalisation operation is adopted to generate the output of the Transformer network. The operation can be expressed as follows:

$$y_{feedforward} = LayerNorm\left[y_{norm} + \left(\max\left(0, y_{norm}W_1 + b_1\right)W_2 + b_2\right)\right] \tag{13}$$

where W_1 and W_2 are the weight of feed-forward layers, b_1 and b_2 are the biases of the feed-forward layers.

Through the process of Transformer, the long-term dependency within the integrated data can be further revealed. Then, the output is flattened and fed into two fully connected layers to predict MEC.

3.3. Evaluation metrics

To verify the effect of physical knowledge and predictive accuracy of the proposed method, this paper uses four metrics, mean square error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE) to evaluate the performance on MEC prediction. The four metrics are listed as follows.

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left(A_t - F_t\right)^2 \tag{14}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|A_t - F_t\right| \tag{15}$$

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{16}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(A_t - F_t\right)^2} \tag{17}$$

where, n is the number of observations, A_t is the actual value and F_t is the predicted value.

4. Experimental setup

4.1. Data collection and PIMT model description

In this paper, equipment information encompasses property information from machine tools, cutting tools, and workpieces, including machine tool models, cutting tool dimensions, and workpiece material etc. Sensor data, including spindle motor power, feed-axis power, and three-directional cutting forces. Process parameters comprise cutting speed, feed rate, and cut depth. Machining condition data, such as workshop temperature and coolant flow rate. To conduct machining operations and gather machining data, an experimental platform for MEC prediction is established in this paper. Illustrated in Figure 4, the platform consists of a machining execution system and a data acquisition system. The machining execution system is utilized for conducting machining operations. The data acquisition system is employed to collect machining data. Specifically, a power tester typed WT1800 is used to measure the power signals of the machine tools, utilizing a sampling frequency of 1 kHz. Additionally, the data of workshop temperature and coolant flow rate are acquired by air conditioning and other equipment. Conversely, the sources of equipment information and process parameter include CAD models, CAM systems, and process manuals. In addition, the synthetic data is fitted from the process parameter. Through this platform and experimental design, a total of 16,665 samples were collected, forming a substantial foundation for training and evaluating the proposed PIMT model.

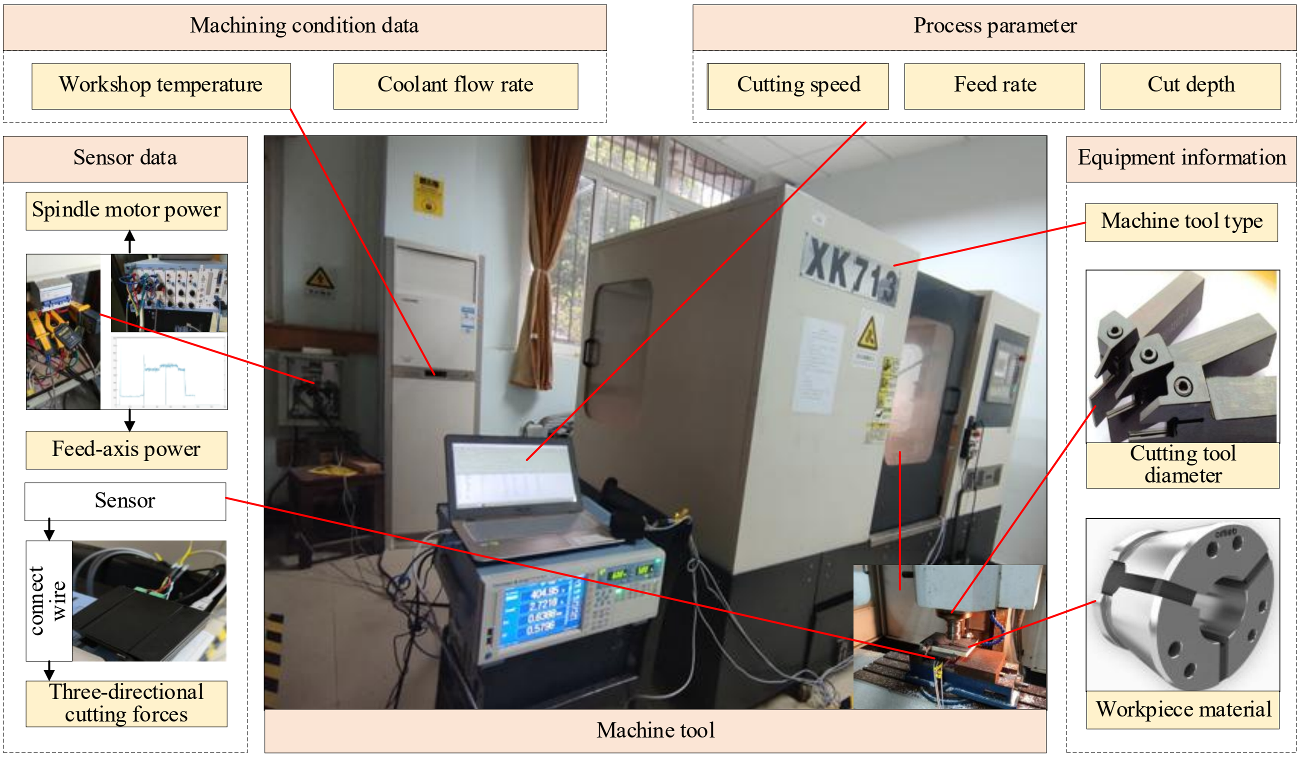

Figure 4. Experimental platform for MEC prediction

As mentioned in Section 3.2, synthetic data, sensor data, process parameter, equipment information, and machining condition data exhibit diverse modalities, encompassing static data, structured data, discrete event sequence data, and continuous time-sequence data. The details of the data are provided in Table 2.

Following data collection, one crucial issue is integrating them into the proposed PIMT model. Parameter configuration holds the potential to influence the performance of the PIMT model. In this study, Batch Normalization was implemented for input multi-modal data, and the vital network parameters are detailed in Table 3. Moreover, determining the training parameters is essential. The chosen optimizer was Adam, with a learning rate of 0.05 and a training epoch count of 100. An early stopping strategy was also implemented to achieve optimal results throughout the training process. The mean square error was selected as the loss function due to its suitability for the regression task.

Table 2. Data description

Classification	Items	Data characteristic
Equipment information	Machine model	static data
	Machine type	static data
	Power Rating	static data
	Workpiece material	static data
	Workpiece geometry	static data
	Tool type	static data
	Tool material	static data
	Tool diameter	static data
	Cutting edges	static data
Sensor data	Spindle motor power	continuous time-sequence data

	Feed-axis power	continuous time-sequence data
	Three-directional cutting forces	continuous time-sequence data
Process parameter	cutting speed	discrete event sequence data
	feed rate	discrete event sequence data
	cut depth	discrete event sequence data
Machining condition data	workshop temperature	structured data
	coolant flow rate	structured data
Synthetic data	*A, B, C*	discrete event sequence data
	K, D, E, F, G	discrete event sequence data

Table 3. Detailed parameter configuration of the PIMT model.

Layer	parameter
Convolutional kernel size	10
Number of 1D-convolutional layers	3
Number of fully connected layers in the sub-network	2
Number of neurons in the hidden layers of the Transformer	100
Number of heads for multi-head self-attention	4
Dropout	0.5

4.2. Comparative Experiments

In this subsection, three comparative experiments are designed and set up to assess the feasibility and superiority of the proposed PIMT model for MEC prediction. The specifics of these experiments are outlined as follows.

(1) Experiment 1 involves utilizing the PIMT model to forecast MEC for two industrial milling machine tools (XK714D, XK713) with three workpiece types (PA66, Al 6061, and 1045 steel) and three cutting tool diameters (8mm, 10mm, and 12mm) to reveal the effects of physical knowledge exploration. The evaluation metrics are applied to assess the outcomes when synthetic data is introduced as an additional input to the PIMT model, and when it is not used. The data partitioning strategy uses an 8:2 split to divide the dataset (with/without synthetic data) into fixed training and test sets.

(2) In Experiment 2, two comparisons between energy consumption predicted values and actual values are established, one between real-time predicted values and actual values, and the other between predicted values and actual values in cutting and air cutting states. These comparisons aim to illustrate how multimodal data fusion enhances the prediction accuracy of energy consumption in MEC. The data partitioning strategy uses an 8:2 split to divide the dataset (including synthetic data) into fixed training and test sets.

(3) Experiment 3 aims to demonstrate the effectiveness of the proposed PIMT model by comparing its predicted results with state-of-the-art benchmarking algorithms, including Bi-long-short term memory network (Bi-LSTM), spatial attention-based convolutional transformer (SAConvFormer), and Merged-long-short term memory network (M-LSTM). The evaluation was conducted using a 10-fold cross-validation strategy to ensure robust performance comparison. It is important to note the architectural differences among the benchmarks. For Bi-LSTM and SAConvFormer, which are purely temporal prediction algorithms, all the variables are considered temporal data and fed into the network. In contrast, for M-LSTM and the proposed PIMT, different types of variables are fed into separate networks. The configuration details of the algorithms are described as follows.

- Bi-LSTM [44]: A neural network that consists of three bi-directional LSTM layers and a fully connected layer, with 1000 nodes in the hidden layers.
- SAConvFormer [45]: A neural network comprising convolutional layers, spatial attention modules, max-pooling layers for sequential feature extraction, and a transformer network for sequential pattern mining, with 1000 nodes in the hidden layers.
- M-LSTM [46]: A neural network consisting of LSTM layers and a fully connected layer, designed to handle both sequential data and ordinary numeric data simultaneously, with 1000 nodes in the hidden layers.

In the above comparative experiments, the Taguchi methods are employed to set the machining parameters during machining, the details are provided in Table 4.

Table 4. Detailed machining parameters of the experiments.

Parameters	Information	Unit
Tool material	PCBN	/
Blade diameter	8/10/12	mm
Workpiece material	PA66/Al 6061/AISI 1045 steel	/
Spindle rotation speeds	600/900/1200/1500/1800	r/min
Feed rates	10/20/30/40/50	mm/min
Cutting depth	0.5/1/1.5/2/2.5	mm

5. Experimental results

5.1. The results of the PIMT model

To demonstrate the performance of the proposed PIMT model, comparison experiments involving Bi-LSTM, SAConvFormer, and M-LSTM are assessed using 10-fold cross-validation. The evaluation metrics employed are MAPE and RMSE, as illustrated in Figure 5.

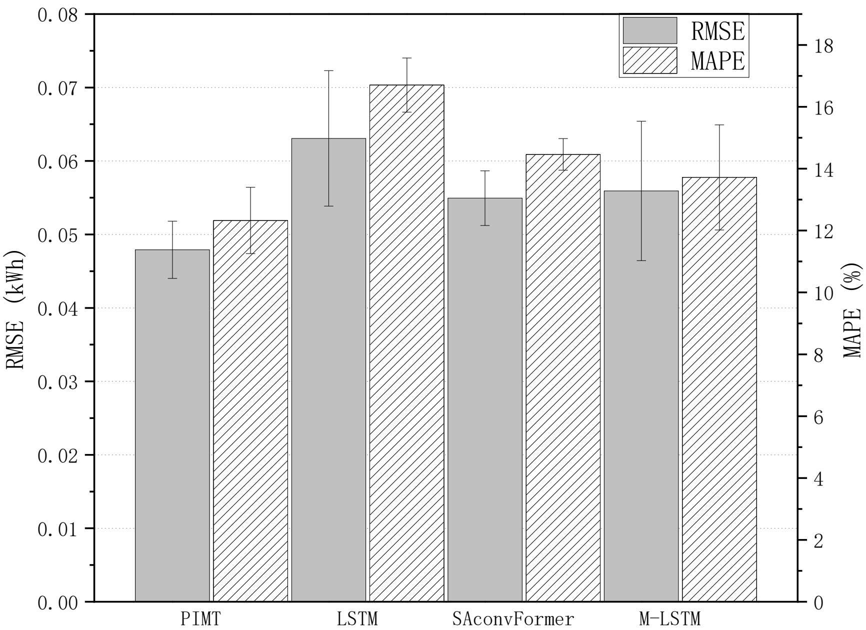

Figure 5. Comparison of algorithm performance in terms of RMSE and MAPE

To further illustrate the performance of the proposed algorithm, the computation load of the benchmarking algorithms was also evaluated in two aspects: (1) the number of trainable parameters, and (2) the training time for 100 epochs. All tests were performed on a computer with an Intel i9-10920X 3.50 GHz CPU and an Nvidia GeForce RTX 3090 graphics card. The results are presented in Figure 6 and Table 5.

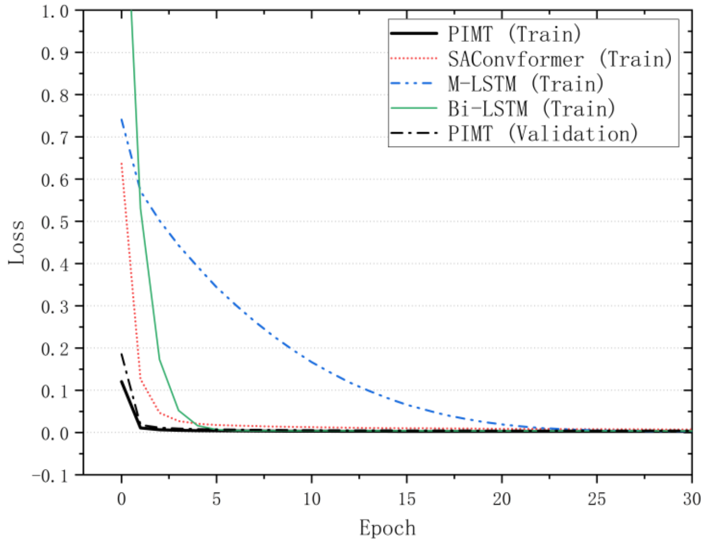

Figure 6. Convergence curves for various algorithms

Table 5. Comparison of algorithm cost

	PIMT-Net	Bi-LSTM	SAConvFormer	M-LSTM
Trainable parameters	**13,964**	27,737	28,689	42,121
Training time (s)	**43**	195	44.7	72.9

5.2. The results of physical knowledge exploration

To illustrate the effect of physical knowledge exploration on MEC prediction, the results are compared with and without synthetic data in Experiment 1. These results encompass two types of machine tools, three workpiece materials, and three cutting tools. The evaluation metrics encompass MSE, MAE, RMSE, and MAPE, with the results depicted in Figures 7-9.

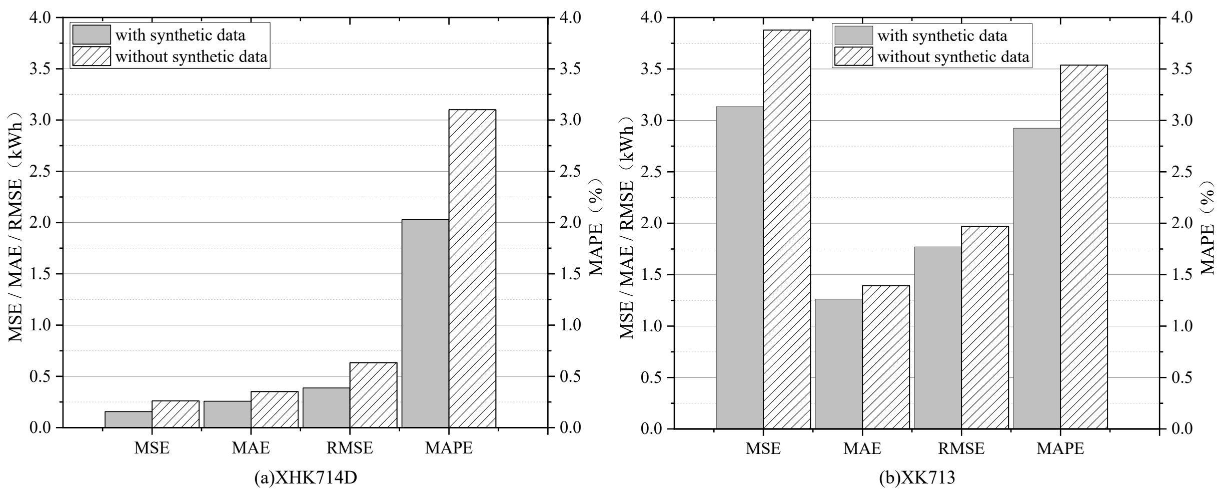

Figure 7. The prediction results of the machine tools

Paper number: JCISE-25-1455. Corresponding author: Xiaolei Liang

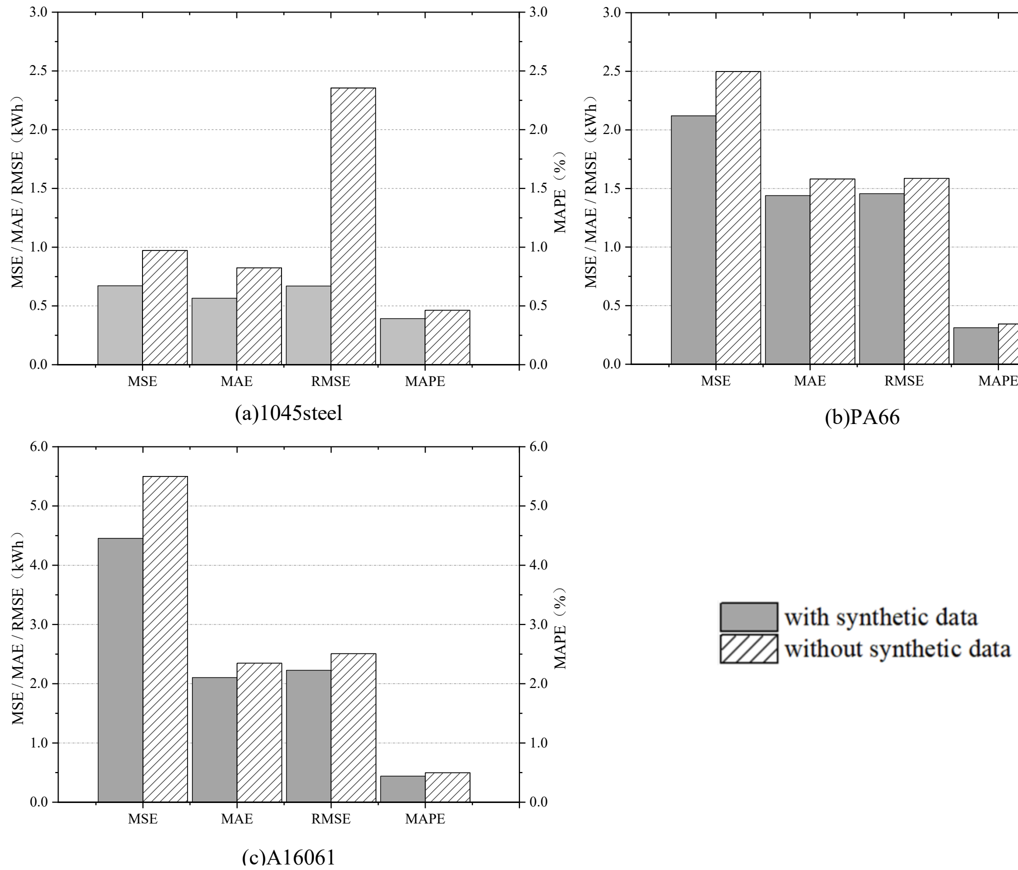

Figure 8. The prediction results of three workpiece materials

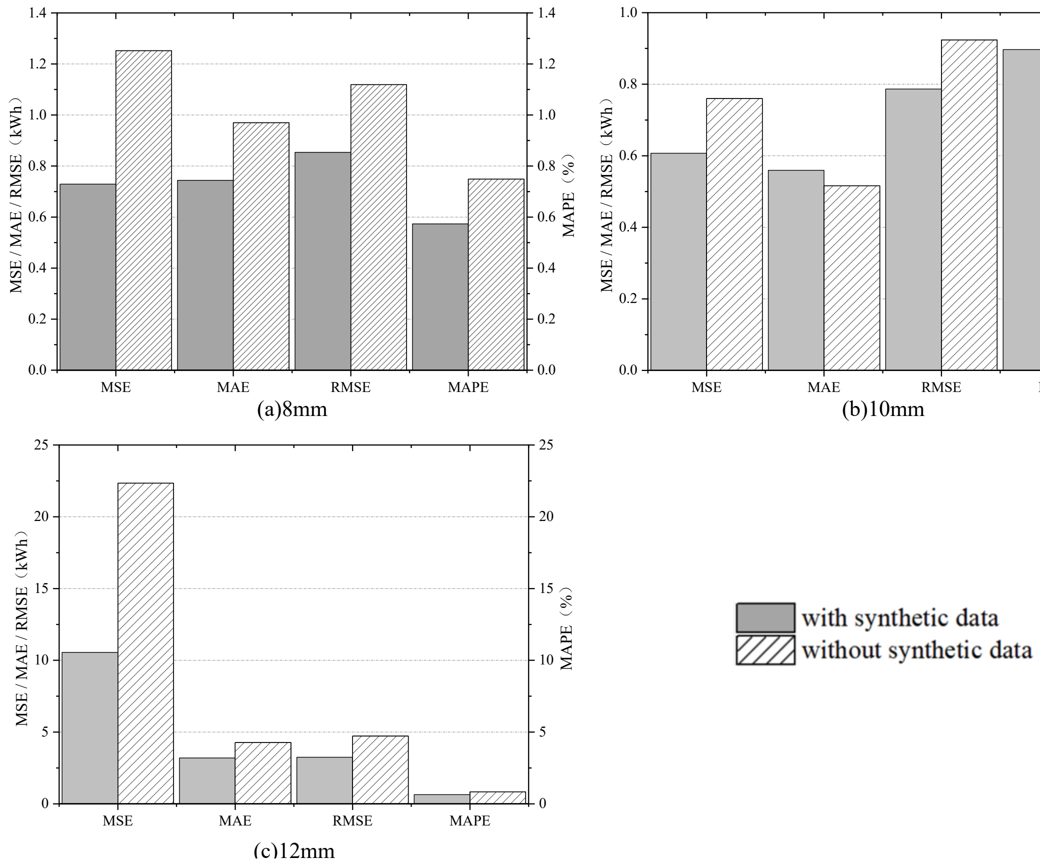

Figure 9. The prediction results of three cutting tools

5.3. The results of multi-modal data fusion

Experiment 2 involves a two-group comparison between predicted values and actual values of MEC to reveal the performance of multi-modal data fusion. The first group validated the effectiveness of the PIMT model for energy consumption prediction by analysing the real-time values of the first 1000 data points under continuous processing conditions and performing cumulative summation on these real-time values (as shown in Figures 10 and 11). In the second group, the energy consumption predicted values for air cutting and cutting were computed and compared with the actual values, as presented in Figure 11.

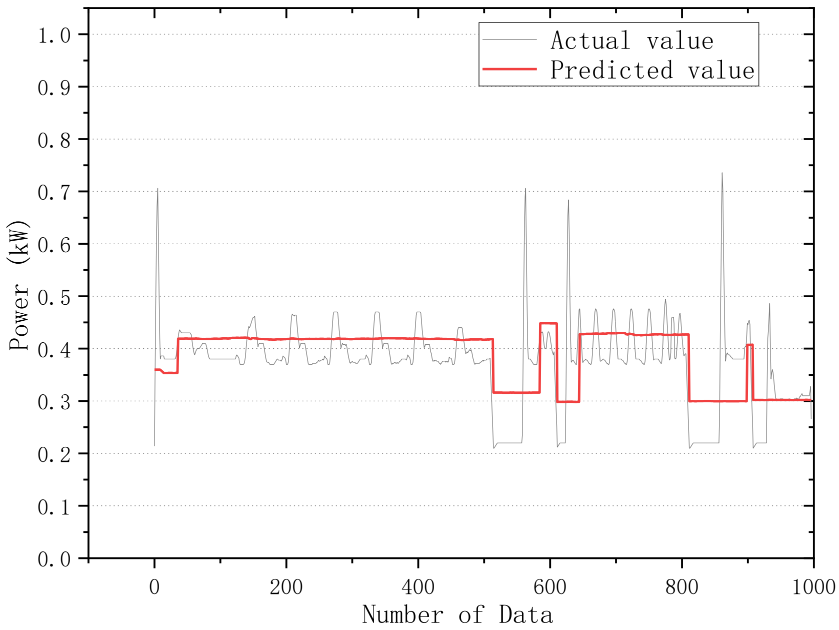

Figure 10. Real-time MEC prediction with the PIMT model

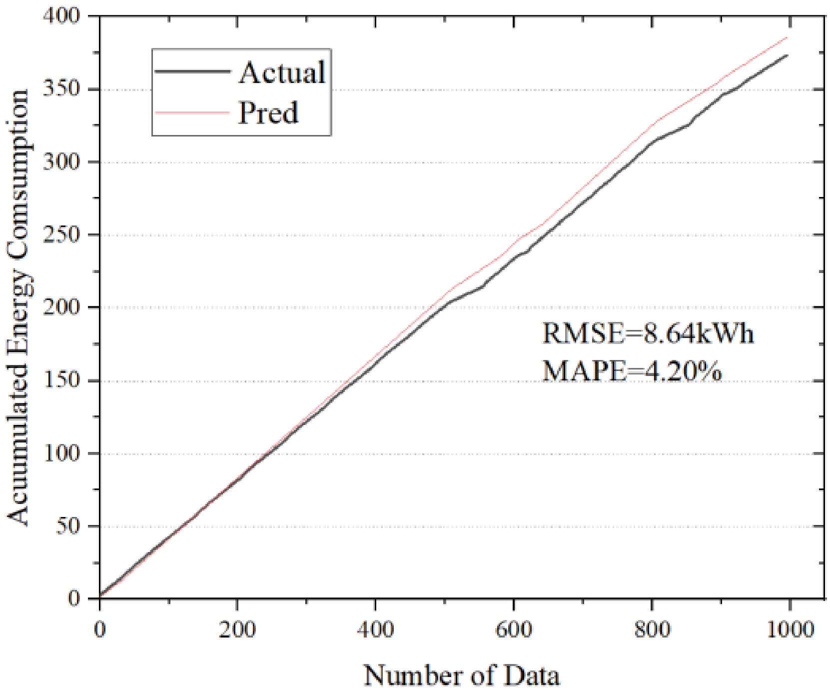

Figure 11. Comparison of accumulated values

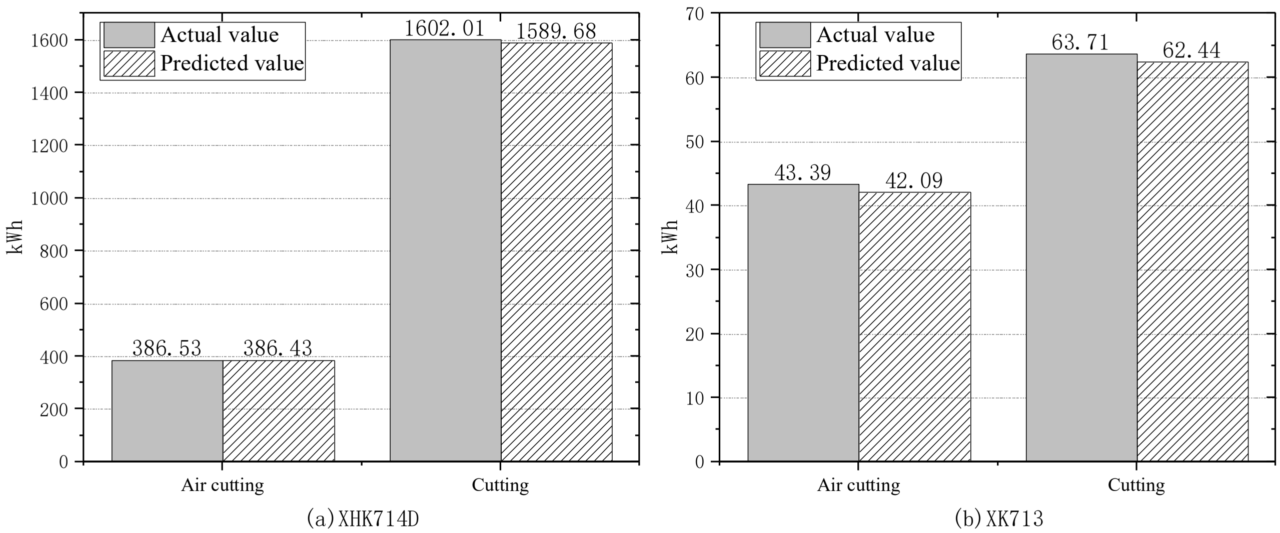

Figure 12. The comparison between actual values and predicted values of different energy consumption

5.4. Discussion

The comparative experimental results unequivocally demonstrate the feasibility and superiority of the proposed PIMT model, which achieves the lowest values in both RMSE (0.048 ± 0.004) and MAPE (12.327% ± 1.1%), as illustrated in Figure 5. More importantly, an in-depth analysis of the error bars reveals distinct advantages in model stability. While Bi-LSTM and M-LSTM exhibit the largest error fluctuations (e.g., RMSE ranges of approx. 0.063 ± 0.009 and 0.056 ± 0.009, respectively), indicating high prediction uncertainty, the proposed PIMT maintains a consistently tight error distribution. Although SAConvFormer shows a comparable fluctuation range to PIMT, its overall performance is inferior, as all the metrics of PIMT lie below those of SAConvFormer. This indicates that PIMT not only achieves higher accuracy on average but also delivers stable and reliable predictions under varying conditions. Therefore, the integration of physical knowledge and multimodal data in PIMT contributes to both enhanced predictive accuracy and improved robustness.

Figure 6 and Table 5 highlight the computational advantages of the PIMT model. When comparing PIMT with SAConvFormer, both of which incorporate transformer networks, the results indicate a significantly superior performance of PIMT in comparison to SAConvFormer. This can be attributed to the complex operational conditions of machining, which make it challenging to systematically describe the energy nature of machine tools through sensor data, process parameter, equipment information, and machining condition data, alone. The findings demonstrate a positive contribution to enhancing the prediction accuracy of MEC through the enrichment of synthetic data. A comparison between PIMT, Bi-LSTM, and M-LSTM demonstrates that the transformer network exhibits more efficient performance. This is attributed to the multi-head self-attention mechanism of the transformer network, which directly captures long-range dependencies for various position combinations in a sequence, significantly enhancing the ability to represent correlations in sequence information. Furthermore, deep learning techniques such as PIMT and M-LSTM outperform Bi-LSTM, which treats all variables as temporal data. In summary, the PIMT not only achieves a lower final loss but also demonstrates significantly faster convergence than other models, all while maintaining a close alignment between training and validation loss, which indicates robust performance without overfitting. And the method exhibits superior performance in MEC modelling and prediction, especially when dealing with diverse types of variables.

As shown in Figures 7–9, the evaluation results obtained using synthetic data are mostly lower than those without synthetic data. For example, based on the four evaluation metrics (MSE, MAE, RMSE, MAPE), the results for XHK714D show relative reductions of 40%, 27%, 39%, and 35%; for 1045 steel, relative reductions of 31%, 9%, 10%, and 17%; and for the 8 mm tool diameter, relative reductions of 42%, 23%, 24%, and 23%. The occasional higher evaluation values observed for the 10 mm tool diameter, when compared to the scenario without synthetic data, are likely attributable to the limited data available for this specific diameter. Despite this particular case, the overall findings demonstrate that incorporating physical knowledge can further improve MEC prediction accuracy. These results also underscore the effective generalization capability of the proposed physical knowledge exploration strategy across MEC prediction scenarios.

Analysing the comparison between predicted and actual values in Figure 10, it is observed that the trend in predicted values largely aligns with that observed under actual machining conditions. Further accumulation of the real-time values indicates minimal discrepancy between the cumulative values derived from actual and predicted scenarios (Figure 11), demonstrating the method's capability to accurately predict the MEC under continuous machining conditions. Furthermore, Figure 12 shows that the predicted values exhibit high correlation with actual values across both energy consumption categories. These findings indicate the robust performance of the proposed method for MEC prediction. It is noteworthy that the energy consumption prediction highest percentage error in the case of XK713 (3.09%) is higher than that in XHK714D (0.78%). This difference can likely be attributed to the smaller data sample size available for XK713. Concurrently, the multi-modal data fusion approach proposed in this paper can maintain diversity and integrity, optimizing the potential of synthetic data, sensor data, process parameter, equipment information, and machining condition data.

To the best of our knowledge, most prevailing DDAs for MEC prediction have depended on algorithm performance and data-efficient integration. This study employs the FCNN and 1D-CNN as tools for efficiently integrating multi-modal data, including synthetic data, sensor data, process parameter, equipment information, and machining condition data. Furthermore, the Transformer network establishes a precise prediction model leveraging its multi-head self-attention mechanism. In comparison with Bi-LSTM, SAConvFormer and M-LSTM, the PIMT effectively resolves the proposed model, achieving superior accuracy while minimizing computation expenses. The results can be adopted by engineers, analysts, and researchers to establish an enhanced energy management framework within the manufacturing industry. However, the multi-modal data fusion approach presented in this study necessitates the creation of tailored deep learning networks based on the data formats, a process predominantly reliant on expert knowledge. Thus, how to establish a unified framework to integrate multi-modal data requires further in-depth research.

6. Conclusions

This study established a PIMT model to address the physical knowledge integration and data fusion limitations in DDAs for MEC prediction. In this model, a synthetic data fitting approach from the empirical equations of MEC is introduced. With this data, it not only enhances the performance of DDAs but also harnesses the latent physical knowledge. The PIMT model, consisting of an FCNN, a 1D-CNN and a Transformer, represents a significant advancement in multimodal data (synthetic data, sensor data, process parameter, equipment information, and machining condition data) fusion. It further enhances the prediction accuracy of the DDAs in MEC prediction. The combination of physical knowledge exploration, multimodal data fusion, and transformative capabilities establishes a promising avenue for future research endeavours within this domain.

While this study has implemented PIML at the data level by incorporating synthetic features, this represents merely an initial step. To further explore the integration of physical principles, our future work will prioritize introducing physics-informed loss functions in PIML. Beyond this, the emergence of powerful generative AI presents significant potential. We plan to leverage these techniques, such as diffusion models, to generate high-fidelity data for rare or costly-to-measure machining scenarios, thereby further enhancing the model's robustness and ability to generalize to unseen conditions. This concrete technical pathway, which progresses from data-level to loss-level physics integration and is augmented by generative AI for data augmentation, will establish a solid foundation for developing more reliable, generalizable, and automation-friendly MEC prediction systems.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 52575594, 51975432), the China Scholarship Council for Visiting Scholars (No. 202008420116), "The 14th Five Year Plan" Hubei Provincial advantaged characteristic disciplines (groups) project of Wuhan University of Science and Technology (No. 2023B0405).

Conflict of Interest

There are no conflicts of interest.

Data availability statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

References

[1] Pimenov, D. Y., Mia, M., Gupta, M. K., Machado, Á. R., Pintaude, G., Unune, D. R., Khanna, N., Khan, A. M., Tomaz, Í., Wojciechowski, S., and Kuntoğlu, M., 2022, "Resource saving by optimization and machining environments for sustainable manufacturing: A review and future prospects," Renew. Sust. Energ. Rev., **166**, p. 112660.

[2] IEA, 2021, "INTERNATIONAL ENERGY AGENCY," Key World Energy Statistics 2021, IEA, Paris, Licence: CC BY 4.0.

[3] Pawanr, S., Garg, G. K., and Routroy, S., 2022, "Prediction of energy consumption of machine tools using multi-gene genetic programming," Mater. Today. Proc., **58**, pp. 135-139.

[4] Feng, C., Wu, Y., Li, W., Qiu, B., Zhang, J., and Xu, X., 2023, "Energy consumption optimisation for machining processes based on numerical control programs," Adv. Eng. Inf., **57**, p. 102101.

[5] He, K. Y., Tang R. Z., Zhang, Z. W., and Sun, W. J., 2016, "Energy Consumption Prediction System of Mechanical Processes Based on Empirical Models and Computer-Aided Manufacturing," J. Comput. Inf. Sci. Eng., **16** (4), p. 041008.

[6] Gutowski, T. G., Branham, M. S., Dahmus, J. B., Jones, A. J., Thiriez, A., and Sekulic, D. P., 2009, "Thermodynamic analysis of resources used in manufacturing processes," Environ. Sci. Technol., **43**(5),

pp. 1584-1590.
[7] Liu, P., Liu, F., & Qiu, H., 2017, “A novel approach for acquiring the real-time energy efficiency of machine tools,” Energy, **121**, pp. 524-532.
[8] Shi, J., Yu, T., Goebel, K., and Wu, D., 2020, “Remaining Useful Life Prediction of Bearings Using Ensemble Learning: The Impact of Diversity in Base Learners and Features,” J. Comput. Inf. Sci. Eng., **21**(2), p. 021004.
[9] He, Y., Wu, P., Li, Y., Wang, Y., Tao, F., and Wang, Y., 2020, “A generic energy prediction model of machine tools using deep learning algorithms,” Appl. Energy, **275**, p. 115402.
[10] Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L., 2021, “Physics-informed machine learning,” Nat. Rev. Phys., **3**(6), pp. 422-440.
[11] Leng, J. W., Zuo, K. W., Xu, C. Y., Zhou, X. L., Zheng, S., Kang, J. W., Liu, Q., Chen, X., Shen, W. M., Wang, L. H., and Gao, R. X., 2025, “Physics-informed machine learning in intelligent manufacturing: a review,” J. Intell. Manuf., pp. 1-43.
[12] Wang, Z., Cui, L., Guo, W., Zhao, L., Yuan, X., Gu, X., Tang, W. Z., Bu, L. G., and Huang, W., 2022, “A design method for an intelligent manufacturing and service system for rehabilitation assistive devices and special groups,” Adv. Eng. Inf., **51**, p. 101504.
[13] Jiang, X., Tian, Z., Liu, W., Suo, Y., Chen, K., Xu, X., and Li, Z., 2022, “Energy-efficient scheduling of flexible job shops with complex processes: A case study for the aerospace industry complex components in China,” J. Ind. Inf. Integr., **27**, p. 100293.
[14] Liu, F., Xu Z. J., and Dan B., 1995, “Energy characteristics of machining systems and its application,” Beijing: China Machine Press.
[15] Mori, M., Fujishima, M., Inamasu, Y., and Oda, Y., 2011, “A study on energy efficiency improvement for machine tools,” CIRP Ann., **60**(1), pp. 145-148.
[16] Balogun, V. A., and Mativenga, P. T., 2013, “Modelling of direct energy requirements in mechanical machining processes,” J. Clean. Prod., **41**, pp. 179-186.
[17] Ma, F., Zhang, H., Cao, H., and Hon, K. K. B., 2017, “An energy consumption optimization strategy for CNC milling,” Int. J. Adv. Manuf. Technol., **90**, pp. 1715-1726.
[18] Liu, Z. Y., Li, C., Fang, X. Y., and Guo, Y. B., 2018, “Cumulative energy demand and environmental impact in sustainable machining of Inconel superalloy,” J. Clean. Prod., **181**, pp. 329-336.
[19] Rossit, D. A., Tohmé, F., and Frutos, M., 2019, “A data-driven scheduling approach to smart manufacturing,” J. Ind. Inf. Integr., **15**, pp. 69-79.
[20] Hu, L., Tang, R., He, K., and Jia, S., 2015, “Estimating machining-related energy consumption of parts at the design phase based on feature technology,” Int. J. Prod. Res., **53**(23), pp. 7016-7033.
[21] Brillinger, M., Wuwer, M., Hadi, M. A., and Haas, F., 2021, “Energy prediction for CNC machining with machine learning,” CIRP Ann. Manuf. Technol., **35**, pp. 715-723.
[22] Bhinge, R., Park, J., Law, K. H., Dornfeld, D. A., Helu, M., and Rachuri, S., 2017, “Toward a generalized energy prediction model for machine tools,” J. Manuf. Sci. Eng., **139**(4), p. 041013
[23] Sheng, Y., Zhang, G., Zhang, Y., Luo, M., Pang, Y., and Wang, Q., 2024, “A multimodal data sensing and feature learning-based self-adaptive hybrid approach for machining quality prediction,” Adv. Eng. Inf., **59**, p. 102324.
[24] Salunkhe, V. G., Khot, S. M., Jadhav, P. S., Yelve, N. P., and Kumbhar, M. B. 2024, “Experimental Investigation Using Robust Deep VMD-ICA and 1D-CNN for Condition Monitoring of Roller Element Bearing,” ASME. J. Comput. Inf. Sci. Eng., **24**(12), p. 124501.
[25] Chen, C., Liu, Y., Kumar, M., Qin, J., and Ren, Y., 2019, “Energy consumption modelling using deep

learning embedded semi-supervised learning," Comput. Ind. Eng., **135**, pp. 757-765.

[26] Qin, J., Liu, Y., and Grosvenor, R., 2018, "Multi-source data analytics for AM energy consumption prediction," Adv. Eng. Inf., **38**, pp. 840-850.

[27] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I., 2017, "Attention is all you need," Adv Neural Inf Process Syst, **30**, pp. 6000–6010.

[28] Chen, C., Liu, C., Wang, T., Zhang, A., Wu, W., and Cheng, L., 2023, "Compound fault diagnosis for industrial robots based on dual-transformer networks," J. Manuf. Syst., **66**, pp. 163-178.

[29] Liu, Z., and Guo, Y., 2018, "A hybrid approach to integrate machine learning and process mechanics for the prediction of specific cutting energy," CIRP Ann., **67**(1), pp. 57-60.

[30] Lv, L., Deng, Z., Yan, C., Liu, T., Wan, L., and Gu, Q., 2020, "Modelling and analysis for processing energy consumption of mechanism and data integrated machine tool," Int. J. Prod. Res., **58**(23), pp. 7078-7093.

[31] Li, Y., Wang, J., Huang, Z., and Gao, R. X., 2022, "Physics-informed meta learning for machining tool wear prediction," J. Manuf. Syst., **62**, pp. 17-27.

[32] Leng, J. W., Su, X. Y., Liu, Z. A., Zhou, L. H., Chen, C., Guo, X., Wang, Y. W., Wang, R., Zhang, C., Liu, Q., Chen, X., Shen, W. M., and Wang, L. H., 2025, "Diffusion model-driven smart design and manufacturing: Prospects and challenges," J. Manuf. Syst., **82**, pp. 561-577.

[33] Gao, J., Li, P., Chen, Z., and Zhang, J., 2020, "A survey on deep learning for multimodal data fusion," Neural Comput., **32**(5), pp. 829-864.

[34] Yang, Z., Baraldi, P., and Zio, E., 2021, "A multi-branch deep neural network model for failure prognostics based on multimodal data," J. Manuf. Syst., **59**, pp. 42-50.

[35] Chen, Y., Rao, M., Feng, K., and Zuo, M. J., 2022, "Physics-Informed LSTM hyperparameters selection for gearbox fault detection," Mech. Syst. Signal Process., **171**, p. 108907.

[36] Qian, M., Sun, S., and Li, X., 2021, "Multimodal Data and Multiscale Kernel-Based Multistream CNN for Fine Classification of a Complex Surface-Mined Area," Remote. Sens., **13**, p. 5052.

[37] Gu, C., Dai, C., Shi, X., Wu, Z., and Chen, C., 2022, "A cloud-based deep learning model in heterogeneous data integration system for lung cancer detection in medical industry 4.0," J. Ind. Inf. Integr., **30**, p. 100386.

[38] Leng, J. W., Zheng, K. Y., Li, R.J., Chen, C., Wang, B. C., Liu, Q., Chen, X., and Shen, W. M., 2026, "AIGC-empowered smart manufacturing: Prospects and challenges," Robot. Comput.-Integr. Manuf., **97**, p. 103076.

[39] Zamiela, C., Stokes, R., Tian, W., Priddy, M. W., and Bian, L., 2025, "Advancing Thermal Physics Informed PointNet Distortion Prediction Capabilities in Wire Arc Directed Energy Deposition," J. Comput. Inf. Sci. Eng., **25**(11), pp. 111003.

[40] Chen, B., Feng, Q., Niu, B., Yan, F., Gao, B., Yang, J., ... and Liu, J., 2022, "Multi-modal fusion of satellite and street-view images for urban village classification based on a dual-branch deep neural network," Int. J. Appl. Earth Obs. Geoinf., **109**, p. 102794.

[41] Mou, L., Zhou, C., Zhao, P., Nakisa, B., Rastgoo, M. N., Jain, R., and Gao, W., 2021, "Driver stress detection via multimodal fusion using attention-based CNN-LSTM," Expert Syst. Appl., **173**, p. 114693.

[42] Mitra, S., Mukhopadhyay, R., and Chattopadhyay, P., 2022, "PSO driven designing of robust and computation efficient 1D-CNN architecture for transmission line fault detection," Expert Syst. Appl., **210**, p. 118178.

[43] Qin, Y., Yang, J., Zhou, J., Pu, H., and Mao, Y., 2023, "A new supervised multi-head self-attention autoencoder for health indicator construction and similarity-based machinery RUL prediction," Adv. Eng. Inf., **56**, p. 101973.

[44] Zhang, J., Wang, P., Yan, R., and Gao, R. X., 2018, "Long short-term memory for machine remaining life prediction," J. Manuf. Syst., **48**, pp. 78-86.
[45] Chen, C., Wang, T., Liu, Y., Cheng, L., and Qin, J., 2022, "Spatial attention-based convolutional transformer for bearing remaining useful life prediction," Meas. Sci. Technol., **33**(11), p. 114001.
[46] Chen, C., Liu, Y., Sun, X., Di Cairano-Gilfedder, C., and Titmus, S., 2021, "An integrated deep learning-based approach for automobile maintenance prediction with GIS data," Reliab. Eng. Sys. Saf., **216**, p. 107919.